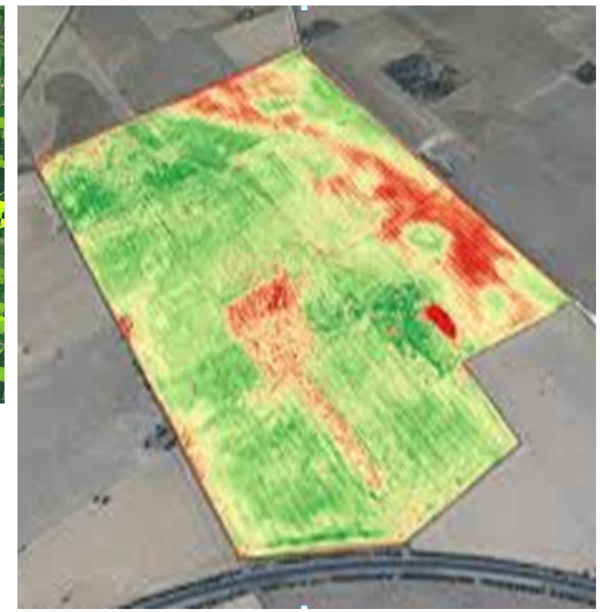
Analyzing satellite data in Stata

Austin Nichols & Hiren Nisar
Abt Associates Inc.

July 27, 2017
Stata Conference
Baltimore, MD



Background

- Recently, there has been an increased interest in remote sensing data for evaluation purposes (especially pictures taken by satellites, but less often drone-based data collection).
  - Kemmerer, Savastano, and Faller (1976) – used Landsat in the 1970's to find fishing grounds.
  - Henderson, Storeygard, and Weil (2012) – estimated GDP growth from satellite images of lights in settled areas
  - Neal, et al. (2016) – demonstrate a method for estimating poverty from high-resolution satellite imagery.
  - Researchers at Abt used satellite images to quantify the spread of oil from Deepwater Horizon (Abt 2016).

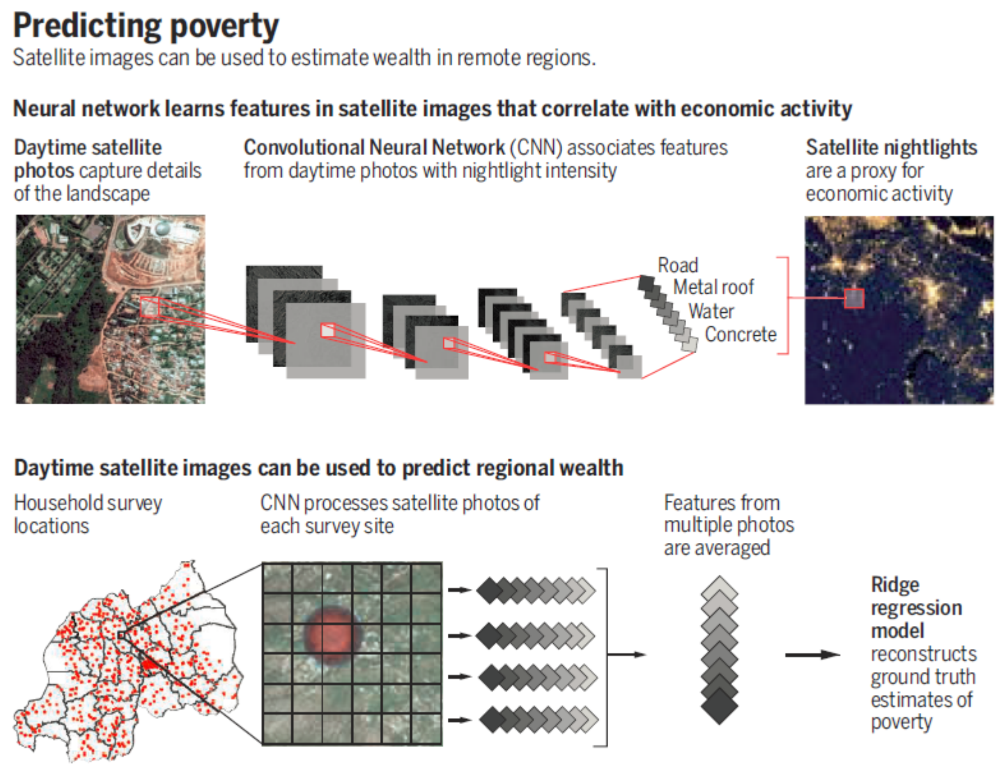
Using Satellite Images in Stata

- `bmp2dta`-user-written command converts a Windows 24-bit bitmap image into a Stata dataset;
- Variables included
  - I and J are rows and columns;
  - R, G, & B are red green and blue content for each pixel;
- Can do standard image processing before converting images to Stata datasets
- Further, can also do custom processing afterwards.

Newer Techniques Available in Stata

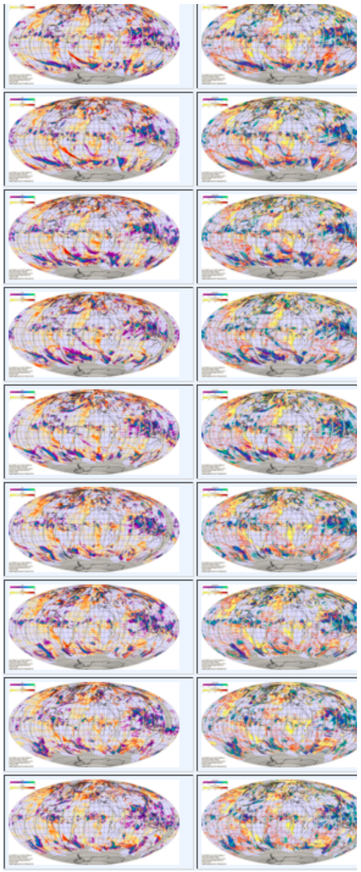
- Can use the suite of Spatial autoregressive (SAR) models new to Stata 15.
- Can also use the full complement of R, Python, & Fortran machine learning tools
  - Or newer tools for Stata (not yet publicly released)

As in Jean et al. (2016):



Many Sources for Satellite Images

- USGS and NASA provide images at high resolution (current and historical)



ASTER Data Products – Suite (n=16)

Shortname	Level	ASTER Product Name	Nominal Resolution (m)	Data Type
AST_L1A	1A	Reconstructed Unprocessed Instrument Data	15, 30, 90	Scene
AST_L1AE	1A	Reconstructed Unprocessed Instrument Data - Expedited	15, 30, 90	Scene
AST_L1BE	1B	Registered Radiance at the Sensor - Expedited	15, 30, 90	Scene
AST_L1B	1B	Registered Radiance at the Sensor	15, 30, 90	Scene
AST_05	2	Surface Emissivity	90	Scene
AST_07	2	Surface Reflectance - VNIR & SWIR	15, 30	Scene
AST_07XT	2	Surface Reflectance - VNIR & Crosstalk Corrected SWIR	15, 30	Scene
AST_08	2	Surface Kinetic Temperature	90	Scene
AST_09	2	Surface Radiance - VNIR & SWIR	15, 30	Scene
AST_09XT	2	Surface Radiance - VNIR & Crosstalk Corrected SWIR	15, 30	Scene
AST_09T	2	Surface Radiance TIR	90	Scene
ASTGTM	3	ASTER Global Digital Elevation Model	30 grid	Tile
AST14DEM	3	Digital Elevation Model	30	Scene
AST14OTH	3	Registered Radiance at the Sensor - Orthorectified	15, 30, 90	Scene
AST14DMO	3	Digital Elevation Model & Registered Radiance at the Sensor - Orthorectified	15, 30, 90	Scene
NAALSED	3	North American ASTER Land Surface Emissivity Database	100	Tile

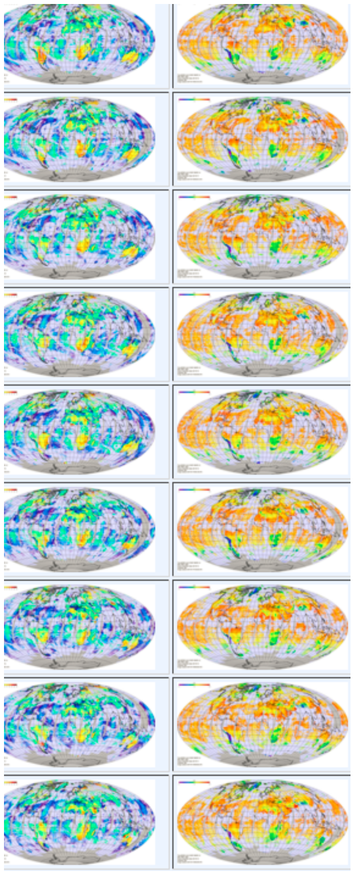
Free (for L1B, U.S. & Territories collection only)

All else $60 for approved Federal Users

Scene dimensions: 60 x 60 km; Tile dimensions: 1 x 1 degree

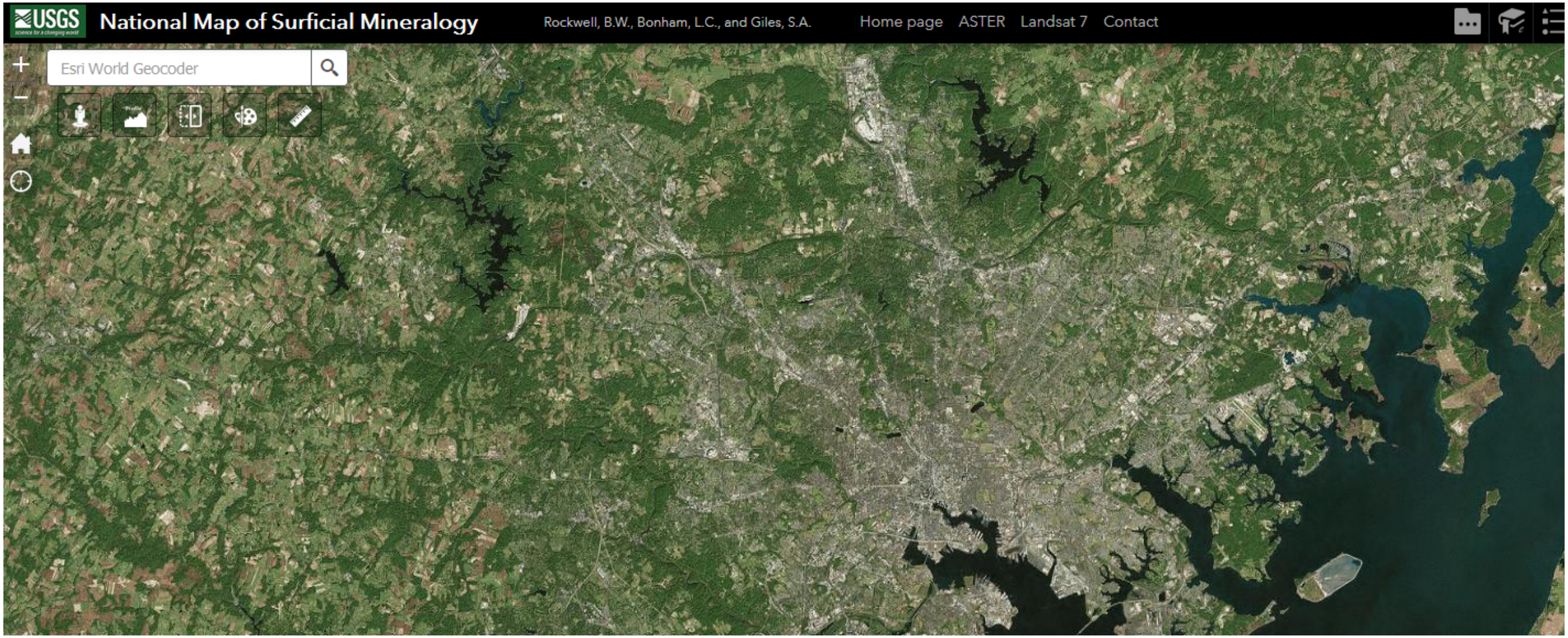
USGS

NASA



Example: Baltimore, MD

- USGS satellite imagery for Baltimore MD: (https://earthexplorer.usgs.gov)



Example: Baltimore, MD

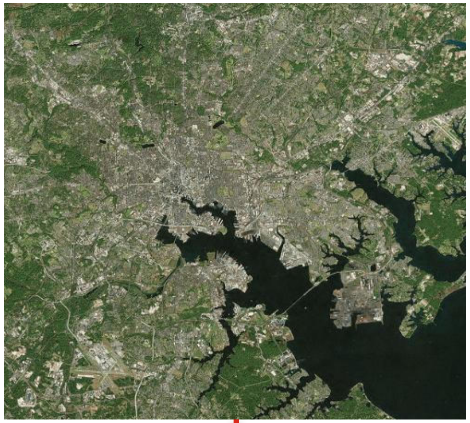
Code :

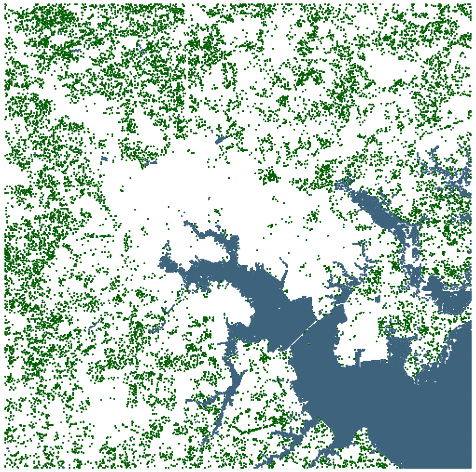
```
bmp2dta using b, pic(balt.bmp)

use b, clear

scatter i j if r<39&g<39&b<39, m(S)
mcol(edkblue)

scatter i j if r<39&g<39&b<39,
msize(vtiny) m(S) mcol(edkblue) ||
scatter i j if g>101 & r<101 & b<100,
msize(vtiny) m(S) mcol(dkgreen)
xsize(5) ysize(5) yla(,nogrid) xsc(off)
ysc(off) leg(off) xla(1/500)
graphr(fc(white) lw(none) margin(zero))
plotr(margin(zero))
```



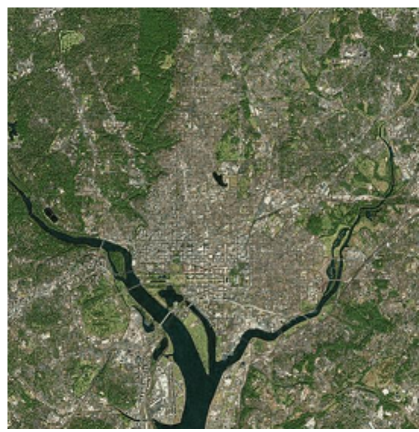


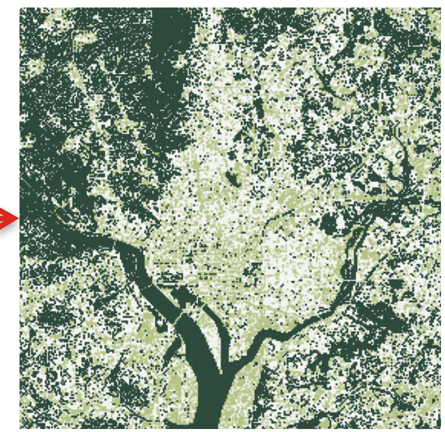
Example: Washington, DC

- Can use more colors, apply filters, do calculations, estimate classifications (clusters).
  - For example, “***cluster kmeans r g b***” can be used to detect pixels dominated by water, trees, or built-up areas.
- Can just plot one class of pixel, to disaggregate layers.

Code:

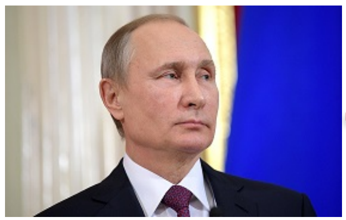
```
bmp2dta using dc276, pic(dc.bmp)
use dc276, clear
cluster kmeans r g b, k(4) gen(c)
forv i=1/4 {
 su r if c==`i', mean
 loc c`i' "`=round(r(mean))'"
 su g if c==`i', mean
 loc c`i' "`c`i'' `=round(r(mean))'"
 su b if c==1, mean
 loc c`i' "`c`i'' `=round(r(mean))'"
 }
sc i j if c==1, msize(small) m(S) mcol("`c1'")||sc i j if c==2, msize(small) m(S) mcol("`c2'%40")||sc i j if c==3, msize(vtiny) m(S) mcol("`c3'%20")||sc i j if c==4, msize(vtiny) m(S) mcol("`c4'%20")  xsize(5) ysize(5) yla(1/200 ,nogrid) xsc(off) ysc(off) leg(off) xla(1/200) graphr(fc(white) margin(zero) lw(none)) plotr(margin(zero)) scale(.2)
```





Example: COPE

- Similarly, can use the same methods to process any picture



Code:

```
bmp2dta using putin, pic(Putin.bmp)
use putin, clear
loc i 1
foreach c in "29 82 97" "86 151 163" "245 255 201" "161 30 34" "97 10 29" {
 g d`i'=(r-real(word("`c'",1)))^2+(g-real(word("`c'",2)))^2+(g-real(word("`c'",3)))^2
 loc i=`i'+1
 }
g color=1 if d1==min(d1,d2,d3,d4,d5)
qui forv i=2/5 {
 replace color=`i' if d`i'==min(d1,d2,d3,d4,d5)
 }
keep if inrange(j,250,1600)
sc i j if color==2, msize(vtiny) m(S) mcol("86 151 163")||sc i j if color==1, msize(vtiny) m(S) mcol("29 82 97")||sc i j if color==3,
msize(vtiny) m(S) mcol("245 255 201")||sc i j if color==4, msize(vtiny) m(S) mcol("161 30 34")||sc i j if color==5, msize(vtiny) m(S)
mcol("97 10 29") xsize(5) ysize(4) yla(1/500,nogrid) xsc(off) ysc(off) leg(off) xla(250/500) graphr(fc(white) margin(zero))
```

References

- Blumenstock, Joshua E. 2016. "Fighting Poverty with data." Science, 353(6301): 753-54.
- Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." American Economic Review, 102(2): 994–1028.
- Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. 2016. "Combining satellite imagery and machine learning to predict poverty." Science, 353(6301): 790-4.
- Kemmerer, Andrew J., Kenneth J. Savastano, and Kenneth Faller. 1976. "LANDSAT Menhaden and Thread Herring Resources Investigation." Slidell, Louisiana: NASA.
- Wallace BP, Brosnan T, McLamb D, Rowles T and others. 2010. "Effects of the *Deepwater Horizon* oil spill on protected marine species." Endangered Species Research, 33:1-7.